REMARKS

I. INTRODUCTION

In response to the Office Action dated July 12, 2004, claims 33, 39, 52, 54, 56, 58, 60, 71, 73, 75, 77 and 79 have been amended. Claims 24-88 remain in the application. Entry of these amendments, and re-consideration of the application, as amended, is requested.

II. CLAIM AMENDMENTS

Applicants' attorney has made amendments to the claims as indicated above. These amendments were made solely for the purpose of clarifying the language of the claims and the numbering of the elements in the claims, and were not required for patentability or to distinguish the claims over the prior art.

III. PRIOR ART REJECTIONS

A. The Office Action Rejections

In paragraphs (10)-(11) of the Office Action, claims 24-34, 36-40, and 42-43 were rejected under 35 U.S.C. §102(e) as being anticipated by Iyer et al., U.S. Patent No. 5,899,992 (Iyer). In paragraphs (13)-(14) of the Office Action, claim 35 was rejected under 35 U.S.C. §103(a) as being unpatentable over Iyer in view of SAS Institute Inc., SAS OnlineDoc®, Version 8, Cary, NC: SAS Institute Inc., (SAS). In paragraphs (15)-(16) of the Office Action, claim 41 was rejected under 35 U.S.C. §103(a) as being unpatentable over Iyer in view of Shafer et al., SPRINT: A Scalable Parallel Classifier for Data Mining, Proceeding of the 22nd VLDB Conference Mumbai, 1996 (Shafer). In paragraphs (17)-(18) of the Office Action, claims 44 and 45 were rejected under 35 U.S.C. §103(a) as being unpatentable over Iyer in view of Bridges, U.S. Patent No. 5,548,770 (Bridges).

Applicants' attorney respectfully traverses these rejections.

B. Applicants' Independent Claims

Applicants' independent claims 24, 44 and 45 are generally directed to computer-implemented system for performing data mining applications. Claim 1 is representative and comprises the elements of:

(a) a computer having one or more data storage devices connected thereto, wherein a relational database is stored on one or more of the data storage devices;

(b) a relational database management system, executed by the computer, for accessing the relational database stored on the data storage devices; and

(c) an analytic application programming interface (API) that generates a set of scalable data mining functions including queries for execution by the relational database management system, executed by the computer, for performing data mining operations directly within the database management system.

C. The Iyer Reference

Iyer describes a method, apparatus, and article of manufacture for a computer implemented scaleable set-oriented classifier. The scalable set-oriented classifier stores set-oriented data as a table in a relational database. The table is comprised of rows having attributes. The scalable set-oriented classifier classifies the rows by building a classification tree. The scalable set-oriented classifier determines a gini index value for each split value of each attribute for each node that can be partitioned in the classification tree. The scalable set-oriented classifier selects an attribute and a split value for each node that can be partitioned based on the determined gini index value corresponding to the split value. Then, the scalable set-oriented classifier grows the classification tree by another level based on the selected attribute and split value for each node. The scalable set-oriented classifier repeats this process until each row of the table has been classified in the classification tree.

D. The SAS Reference

SAS describes a correlation matrix. The correlation matrix table contains Pearson product-moment correlations of Y variables. Correlation measures the strength of the linear relationship between two variables.

E. The Shafer Reference

Shafer describes a scalable parallel classifier for data mining. A decision-tree-based classification algorithm, called SPRINT, removes all memory restrictions and is fast and scalable.

F. The Bridges Reference

Bridges describes an indexing system and method for improving retrieval of data based on a query from a user from a database management system including a main computer and a memory coupled to the main computer for storing the data. The indexing system includes a parallel computer

coupled to the main computer and a parallel disk array coupled to the parallel computer. The invention includes the steps of storing record based data in the memory of the database management system, storing a value based index of selected attributes related to the record based data on the parallel disk array, and determining whether the parallel computer can be used to execute a query to obtain at least a partial result to the query. If so, the query is sent to the parallel computer and the query is executed on the parallel computer to obtain at least a partial result. If a final result cannot be determined on the parallel computer, the partial result from the parallel computer is sent to the database management system and a final result is determined on the database management system using the partial result received from the parallel computer.

G. Applicants' Independent Claims Are Is Patentable Over The Reference

Applicants' attorney respectfully submits that Applicants' independent claims are patentable over the cited references, because the references do not teach or suggest the specific combination of elements recited in Applicants' independent claims.

Specifically, the references do not teach or suggest "an analytic application programming interface (API) that generates a set of scalable data mining functions including queries for execution by the relational database management system, executed by the computer, for performing data mining operations directly within the database management system."

The Office Action asserts that Iyer teaches these limitations of Applicants' claims at col. 3, line 50 through col. 4, line 26 (actually col. 4, line 35), which is set forth below:

> The scalable set-oriented classifier 114 of the present invention resorts to proven scalable database technology to provide a generic solution to the classification problem of scalability. The present invention provides a scalable model for classifying rows of a table within a classification tree. The scalable set-oriented classifier 114 is called the Scalable Supervised Learning Irregardless of Memory (SLIM) Classifier 114. Not only is the SLIM classifier 114 scalable in regions where recently published classifiers are not, but by virtue of building on well known set-oriented database management system (DBMS) primitives, the SLIM classifier 114 instantly exploits several decades of database research and development. The present invention rephrases classification, a data mining method, into analysis of data in a star schema, formalizing further the interrelationship between data mining and data warehousing.
>
> A description of a prototype built using IBM's DB2 product as the RDBMS 108, and experimental results for the prototype are discussed below. Generally, the experimental results indicate that the DB2-based SLIM classifier 114 has desirable properties associating it with linear scalability.

The SLIM classifier 114 is built based on a set-oriented access to data paradigm. The SLIM classifier 114 uses Structured Query Language (SQL), offered by most commercial RDBMS 108 vendors, as the basis for the method. The SLIM classifier 114 is based on well known database methodologies and lets the RDBMS 108 automatically handle scalability. As a result, the SLIM classifier 114 will scale as long as the database scales.

The SLIM classifier 114 leverages the Structured Query Language (SQL) Application Programming Interface (API) of the RDBMS 108, which exploits the benefits of many years research and development pertaining to:

(1) scalability
(2) memory hierarchy
(3) parallelism ([18])
(4) optimization of the executions([16])
(5) platform independence
(6) client server API ([17]).

See S. Sarawagi, Query Processing in Tertiary Memory Databases, VLDB 1995, [hereinafter Sarawag]; S. Sarawagi and M. Stonebraker, Benefits of Reordering Execution in Tertiary Memory Databases, VLDB 1996, [[hereinafter Stonebraker]; G. Bhargava, P. Goel, and B. Iyer, Hypergraph Based Reordering of Outer Join Queries with Complex Predicates, SIGMOD 1995, [hereinafter Bhargava]; T. Nguyen and V. Srinivasan, Accessing Relational Databases from the World Wide Web, SIGMOD 1996, [hereinafter Goel]; C. K. Baru et. al., DB2 Parallel Edition, IBM Systems Journal, Vol. 34, No 2, 1995, [hereinafter Baru]; each of which is which is incorporated by reference herein.

Applicants' attorney disagrees with this analysis.

The only API discussed in the above portions of Iyer is the Structured Query Language (SQL) Application Programming Interface (API) of the relational database management system (RDBMS). However, nothing in this description teaches or suggests "an analytic application programming interface (API) that generates a set of scalable data mining functions including queries for execution by the relational database management system, executed by the computer, for performing data mining operations directly within the database management system." Instead, this API of the RDBMS in Iyer only invokes functions of the RDBMS, but says nothing about generating a set of scalable data mining functions as recited in Applicants' claims. Moreover, the scalable set-oriented classifier of Iyer is not analogous to Applicants' claimed analytic API.

Indeed, the above portions of Iyer do not provide a proper basis for rejecting claims 24-45, because nowhere is Iyer properly applied to the limitations of claims 24-45. Instead, the Office Action relies on general conclusory statements regarding Iyer to reject Applicants' claims, without addressing the specific limitations of those claims or the specific teachings of Iyer.

Further, SAS, Shafer and Bridges fail to overcome the deficiencies of Iyer. Recall that SAS was cited only for Pearson-Product moment correlation and co-variance matrices, Shafer was cited only for performing a split and Bridges was cited only for a parallel computer system.

Moreover, Applicants' claimed invention provides operational advantages over the system disclosed in the various references. Moreover, Applicants' claimed invention solves problems not recognized by the cited references.

Thus, Applicants' attorney submits that independent claims 24, 44, and 45 are allowable over Iyer, SAS, Shafer, and Bridges. Further, dependent claims 25-43 and 46-83 are submitted to be allowable over Iyer, SAS, Shafer, and Bridges in the same manner, because they are dependent on independent claims 24, 44, and 45, respectively, and thus contain all the limitations of the independent claims.

H. <u>Applicants' Dependent Claims Are Is Patentable Over The Reference</u>

Dependent claims 25-43 and 46-83 recite additional novel elements not shown by Iyer, SAS, Shafer, and Bridges.

With regard to dependent claims 25, 46 and 65, which recite that the computer comprises a parallel processing computer comprised of a plurality of nodes, and each node executes one or more threads of the relational database management system to provide parallelism in the data mining operations, the Office Action asserts that these limitations are taught by Iyer. Applicants' attorney disagrees. At the indicated location, Iyer refers to nodes of a classification tree, not the nodes of a parallel processing computer that executes threads of a relational database management system.

With regard to dependent claims 26, 47 and 66, which recite that the scalable data mining functions process data collections stored in the relational database and produce results that are stored in the relational database, these claims stand or fall with claims 24, 44 and 45, respectively.

With regard to dependent claims 27, 48 and 67, which recite that the scalable data mining functions are created by parameterizing and instantiating the analytic API, the Office Action asserts that these limitations are taught by Iyer. Applicants' attorney disagrees. At the indicated location, Iyer refers to an SQL API, not an analytic API for creating scalable data mining functions.

With regard to dependent claims 28, 49 and 68, which recite that the scalable data mining functions are dynamically generated queries comprised of combined phrases with substituting values therein based on parameters supplied to the analytic API, the Office Action asserts that these limitations are taught by Iyer. Applicants' attorney disagrees. At the indicated location, Iyer refers to

an SQL API, not an analytic API for creating scalable data mining functions, and thus does not dynamically generate queries comprised of combined phrases with substituting values therein based on parameters supplied to the analytic API.

With regard to dependent claims 29, 50 and 69, which recite that the scalable data mining functions comprise Data Description functions, Data Derivation functions, Data Reduction functions, Data Reorganization functions, Data Sampling functions, or Data Partitioning functions, these claims stand or fall with claims 28, 49 and 68, respectively.

With regard to dependent claims 30, 51 and 70, which recite that the Data Description functions comprise descriptive statistical functions, these claims stand or fall with claims 29, 50 and 69, respectively.

With regard to dependent claims 31, 52 and 71, which recite that the Data Description functions comprise:

(1) descriptive statistics for one or more numeric columns, wherein the statistics are selected from a group comprising count, minimum, maximum, mean, standard deviation, standard mean error, variance, coefficient of variance, skewness, kurtosis, uncorrected sum of squares, corrected sum of squares, and quantiles,

(2) a count of values for a column,

(3) a calculated modality for a column,

(4) one or more bin numeric columns of counts with overlay and statistics options,

(5) one or more automatically sub-binned numeric columns giving additional counts and isolated frequently occurring individual values,

(6) a computed frequency of one or more column values,

(7) a computed frequency of values for pairs of columns in a column list,

(8) a Pearson Product-Moment Correlation matrix,

(9) a Covariance matrix,

(10) a sum of squares and cross-products matrix, or

(11) a count of overlapping column values in one or more combinations of tables,

these claims stand or fall with claims 29, 50 and 69, respectively.

With regard to dependent claims 32, 53 and 72, which recite that the Data Derivation functions provide column derivations or transformations, the Office Action asserts that these

limitations are taught by Iyer. Applicants' attorney disagrees. At the indicated location, Iyer merely describes transforming table names, but says nothing about column derivations or transformations.

With regard to dependent claims 33, 54 and 73, which recite that the Data Derivation functions comprise:

(1) a derived binned numeric column wherein a new column is bin number,

(2) a n-valued categorical column dummy-coded into "n" 0/1 values,

(3) a n-valued categorical column recoded into n or less new values,

(4) one or more numeric columns scaled via range transformation,

(5) one or more columns scaled to a z-score that is a number of standard deviations from a mean,

(6) one or more numeric columns scaled via a sigmoidal transformation function,

(7) one or more numeric columns scaled via a base 10 logarithm function,

(8) one or more numeric columns scaled via a natural logarithm function,

(9) one or more numeric columns scaled via an exponential function,

(10) one or more numeric columns raised to a specified power,

(11) one or more numeric columns derived via user defined transformation function,

(12) one or more new columns derived by ranking one or more columns or expressions based on order,

(13) one or more new columns derived with quantile 0 to n-1 based on order and n,

(14) a cumulative sum of a value expression based on a sort expression,

(15) a moving average of a value expression based on a width and order,

(16) a moving sum of a value expression based on a width and order,

(17) a moving difference of a value expression based on a width and order,

(18) a moving linear regression value derived from an expression, width, and order,

(19) a multiple account/product ownership bitmap,

(20) a product ownership bitmap over multiple time periods,

(21) one or more counts, amount, percentage means and intensities derived from a transaction summary,

(22) one or more variabilities derived from transaction summary data,

(23) one or more derived trigonometric values and their inverses, including sin, arcsin, cos, arccos, csc, arccsc, sec, arcsec, tan, arctan, cot, and arccot, or

(24) one or more derived hyperbolic values and their inverses, including sinh, arcsinh, cosh, arccosh, csch, arccsch, sech, arcsech, tanh, arctanh, coth, and arccoth,

the Office Action asserts that the limitations of element (12) are taught by Iyer. Applicants' attorney disagrees. At the indicated location, Iyer merely describes forming groupings, not new columns derived by ranking one or more columns or expressions based on order.

With regard to dependent claims 34, 55 and 74, which recite that the Data Reduction functions provide matrix building operations to reduce the amount of data required for analytic algorithms, the Office Action asserts that these limitations are taught by Iyer. Applicants' attorney disagrees. At the indicated location, Iyer merely describes the use of a "count matrix", but says nothing about matrix building operations that reduce the amount of data required for analytic algorithms.

With regard to dependent claims 35, 56 and 75, which recite that the Data Reduction functions comprise:

(1) build one or more data reduction matrices from a group comprising: (i) a Pearson-Product Moment Correlations matrix; (ii) a Covariances matrix; and (iii) a Sum of Squares and Cross Products (SSCP) matrix,

(2) export a resultant matrix, or

(3) restart a matrix operation,

these claims stand or fall with claims 29, 50 and 69, respectively.

With regard to dependent claims 36, 57 and 76, which recite that the Data Reorganization functions provide an ability to reorganize data by joining or de-normalizing pre-processed results into a wide analytic data set, these claims stand or fall with claims 29, 50 and 69, respectively.

With regard to dependent claims 37, 58 and 77, which recite that the Data Reorganization functions comprise:

(1) create a de-normalized new table by removing one or more key columns, or

(2) join a plurality of tables or views into a combined result table,

these claims stand or fall with claims 29, 50 and 69, respectively.

With regard to dependent claims 38, 59 and 78, which recite that the Data Sampling function provides an ability to construct a new table containing a randomly selected subset of the rows in an existing table or view, these claims stand or fall with claims 29, 50 and 69, respectively.

With regard to dependent claims 39, 60 and 79, which recite that the Data Sample function selects one or more data samples of specified sizes from a table these claims stand or fall with claims 29, 50 and 69, respectively.

With regard to dependent claims 40, 61 and 80, which recite that the Data Partitioning function provides an ability to construct a new table containing at least one randomly selected subset of the rows in an existing table or view, wherein the subsets are mutually distinct but all-inclusive subsets of data, these claims stand or fall with claims 29, 50 and 69, respectively.

With regard to dependent claims 41, 62 and 81, which recite that the Data Partitioning function selects one or more data partitions from a table using a database internal hashing technique, these claims stand or fall with claims 29, 50 and 69, respectively.

With regard to dependent claims 42, 63 and 82, which recite that results of the data mining operations are stored in the relational databases, these claims stand or fall with claims 24, 44 and 45, respectively.

With regard to dependent claims 43, 64 and 83, which recite that the relational database management system further comprises an analytical logical data model that stores metadata and processing results from the Scalable Data Mining Functions, the Office Action asserts that these limitations are taught by Iyer. Applicants' attorney disagrees. At the indicated location, Iyer says nothing about an analytical logical data model that stores metadata and processing results from the scalable data mining funcions, but instead merely refers to a training set and leaf node list.

IV. CONCLUSION

In view of the above, it is submitted that this application is now in good order for allowance and such allowance is respectfully solicited.

Should the Examiner believe minor matters still remain that can be resolved in a telephone interview, the Examiner is urged to call Applicants' undersigned attorney.

Respectfully submitted,

GATES & COOPER LLP
Attorneys for Applicants

Howard Hughes Center
6701 Center Drive West, Suite 1050
Los Angeles, California 90045
(310) 641-8797

Date: September 13, 2004

By: [signature]
Name: George H. Gates
Reg. No.: 33,500

GHG/amb